

Darwinian Behavior in a Cold, Sporadically Fed Pool of Ribonucleotides

Michael Yarus

Abstract

A testable, explicit origin for Darwinian behavior, feasible on a chaotic early Earth, would aid origins discussion. Here I show that a pool receiving unreliable supplies of unstable ribonucleotide precursors can recurrently fill this role. By using numerical integration, the differential equations governing a sporadically fed pool are solved, yielding quantitative constraints for the proliferation of molecules that also have a chemical phenotype. For example, templated triphosphate nucleotide joining is $>10^4$ too slow, suggesting that a group more reactive than pyrophosphate activated primordial nucleotides. However, measured literature rates are sufficient if the Initial Darwinian Ancestor (IDA) resembles a 5'-5' cofactor-like dinucleotide RNA, synthesized via activation with a phosphorimidazolid-like group. A sporadically fed pool offers unforeseen advantages; for example, the pool hosts a novel replicator which is predominantly unpaired, even though it replicates. Such free template is optimized for effective selection during its replication. Pool nucleotides are also subject to a broadly based selection that impels the population toward replication, effective selection, and Darwinian behavior. Such a primordial pool may have left detectable modern traces. A sporadically fed ribonucleotide pool also fits a recognizable early Earth environment, has recognizable modern descendants, and suits the early shape of the phylogenetic tree of Earthly life. Finally, analysis points to particular data now needed to refine the hypothesis. Accordingly, a kinetically explicit chemical hypothesis for a terran IDA can be justified, and informative experiments seem readily accessible. Key Words: Cofactor—RNA—Origin of life—Replication—Initial Darwinian Ancestor (IDA). *Astrobiology* 12, 870–883.

Introduction

WHEN A MOLECULAR biologist says “life,” she or he means a system capable of Darwinian evolution, because without such change, there can be no net progress toward the known Tree of Life. Below I discuss the appearance of an Initial Darwinian Ancestor (IDA; Yarus, 2011a), the earliest, simplest system capable of chemically influencing its surroundings (having a phenotype) and capable of its own replication (having a genotype). Because Earth’s present life has a single origin, as realized by Darwin (Darwin, 1859) and now massively confirmed by sequence data (Pace *et al.*, 2012), the historical origin of life need not be likely but must be uniquely accessible. Because such events are more complex and thus arguably occur later, I do not take up compartmentalization (Szostak *et al.*, 2001) or other possible early events.

The nature of the pool

A sporadically fed pool gets occasional supplies of partly activated ribonucleotides from geochemical sources. Is there

a chemically plausible route to a long-lived system that both has a phenotype and a genotype?

The modeled route to this goal (Fig. 1) combines several straightforward ideas. Firstly, I have previously emphasized the value of molecular simplicity in making a structure accessible to primitive synthesis (Illangasekare and Yarus, 1999; Kennedy *et al.*, 2008; Yarus, 2011b). Aminoacyl-RNA synthesis via a 5 nt ribozyme is a characterized experimental example (Yarus, 2011b); such a molecule would occur via untemplated RNA synthesis as soon as mildly activated nucleotides existed. This aminoacyl transfer center would also recur in virtually any RNA synthesized (Illangasekare and Yarus, 2012). The molecular target in Fig. 1 is yet simpler; it is the posited (Yarus, 2011a) simplest molecule with varied biochemical effects which also might replicate; a 5'-5' linked cofactor-like ribodinucleotide. Further, the two nucleotide components of the A-B ribodinucleotide are complementary by hypothesis, simplifying discussion because we need consider a single template and chemically active product. Perhaps A is an adenosine nucleotide and B is a modified chemically

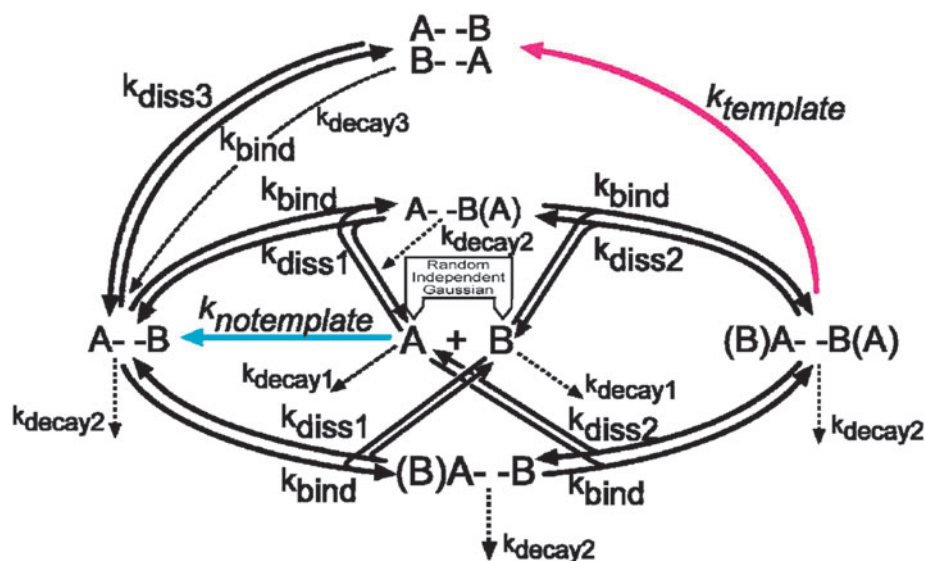


FIG. 1. The reaction scheme for synthesis of a chemically competent, replicating dimer ribonucleotide from randomly timed, Gaussian supplies of precursors—an IDA from a sporadically fed pool. The magenta process (templated dimer synthesis) and the blue process (untemplated dimer synthesis), as well as the peak production of all forms of dimer, Σ A--B, are emphasized in the discussion.

reactive, complementary pyrimidine or pyrimidine-like moiety, together akin to modern NAD (Yarus, 2011a).

After initial untemplated synthesis of A--B (Fig. 1, blue), a product dimer can pair with complementary nucleotides and template its own replication (magenta). Note that because a 5'-5' backbone is symmetrical (does not show polarity), A--B and B--A are identical molecules. All forms of the A, B, and A--B are unstable, including paired (A--B)(B--A) dimer, and disappear at characteristic rates.

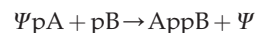
The dual hyphen between A and B commemorates the 5'-5'-linked backbone of the cofactor hypothesis (Yarus, 2011a), but because the analysis below is kinetic rather than structural, the reader is free to think of the outcome with a 3'-5' RNA backbone. Primordial backbone polarities are in any case presently ambiguous (discussed in Yarus, 2011a) and must be decided by Bayesian analysis of current oligonucleotide experiments (explained in Yarus *et al.*, 2005; see also *Discussion*). However, A--B as a 5'-5' cofactor-like dimer offers multiple advantages—for example, a stable alternative RNA backbone. For example, replication is arguably self-limiting because of the unusual $\leftarrow \rightarrow$ backbone symmetry in both template and product. Finally, even in the absence of any knowledge of the evolutionary route to the present, similar structures exist abundantly today as mRNA caps (Schoenberg and Maquat, 2009) and protein cofactors (Yarus, 2011a), providing empirical evidence for the ancient origins and evolutionary potential of the 5'-5' ribodinucleotide.

Calculations below emphasize results that seem sufficiently robust to survive varied assumptions. For this purpose, some reactions take slow rates (see **Calculations** and *Discussion*), so as to underestimate the outcomes of chemical events. In addition, the rarity of reactants is taken into account; substrates are not only unreliably available but evanescent even when they occur. Finally, the fate of the pool is examined over a range, so conclusions are not tied to unique assumed initial conditions.

Separate supplies of the precursors

In Fig. 1, nucleotides A and B appear in the pool at times that are random and uncorrelated. If A and B are purine- and pyrimidine-like, then because likely chemical routes to purine nucleobases (Oro, 1961) and pyrimidine nucleotides (Powner *et al.*, 2009) currently appear chemically distinct, it is plausible that their arrivals would be uncorrelated, because they are expected to be products of differing chemical environments. In addition, A and B quantities are each the sum of several random processes with their own particular distribution of yields. Therefore the Central Limit Theorem suggests that their summed amounts should be distributed as a Gaussian, or normal, distribution (Feller, 1957). Therefore substrates A and B appear as spikes in the pool at random times, with both spike heights distributed as Gaussians.

The dimer synthesis reaction is assumed to be



where Ψ is an activating group for the 5' phosphate of A. For example, if ΨpA is the phosphorimidazolide introduced by the Orgel lab as a model-activated nucleotide (Inoue and Orgel, 1983), this overall reaction is well known (Kanavarioti *et al.*, 1992). In addition, the phosphorimidazolide of A is one of many P-N-bond activated nucleotides that has a possibly prebiotic, moderately efficient synthesis from AMP, imidazole (or a variety of primary amines), and trimetaphosphate (Lohrmann, 1977). Notably, nucleotide pB enters the pool as its unactivated 5' nucleotide, perhaps occurring in several related forms. Thus, we require only one activated nucleotide for all such syntheses, ΨpA . Accordingly, activation chemistry for pA phosphate can be geographically separated from synthesis of the chemically reactive pB, decreasing destructive cross reactions. Possible side reactions in which activated pA reacts with itself are neglected with respect to the above reaction.

Random arrival of substrates

Random appearance means that substrates are completely uncoordinated. Thus, little or no reaction occurs subsequent to many substrate spikes (Fig. 2). Accordingly, any arrival schedule that partially synchronizes A and B arrival would increase their reaction and increase the productivity of the pool. This consideration returns in discussion below.

Universal decay

All biomolecules have finite stabilities, significant on the likely pool timescale. Therefore, A and B nucleotides decay at a first-order rate $k_{\text{decay}1}$ and have mean lifetime $k_{\text{decay}1}^{-1}$ if they do not react. Such rates of disappearance of A and B are usually assumed the same for simplicity, though A might plausibly be taken as the less stable (see below). If A and B do react directly, via $k_{\text{notemplate}}$ (leftward blue arrow in Fig.1;

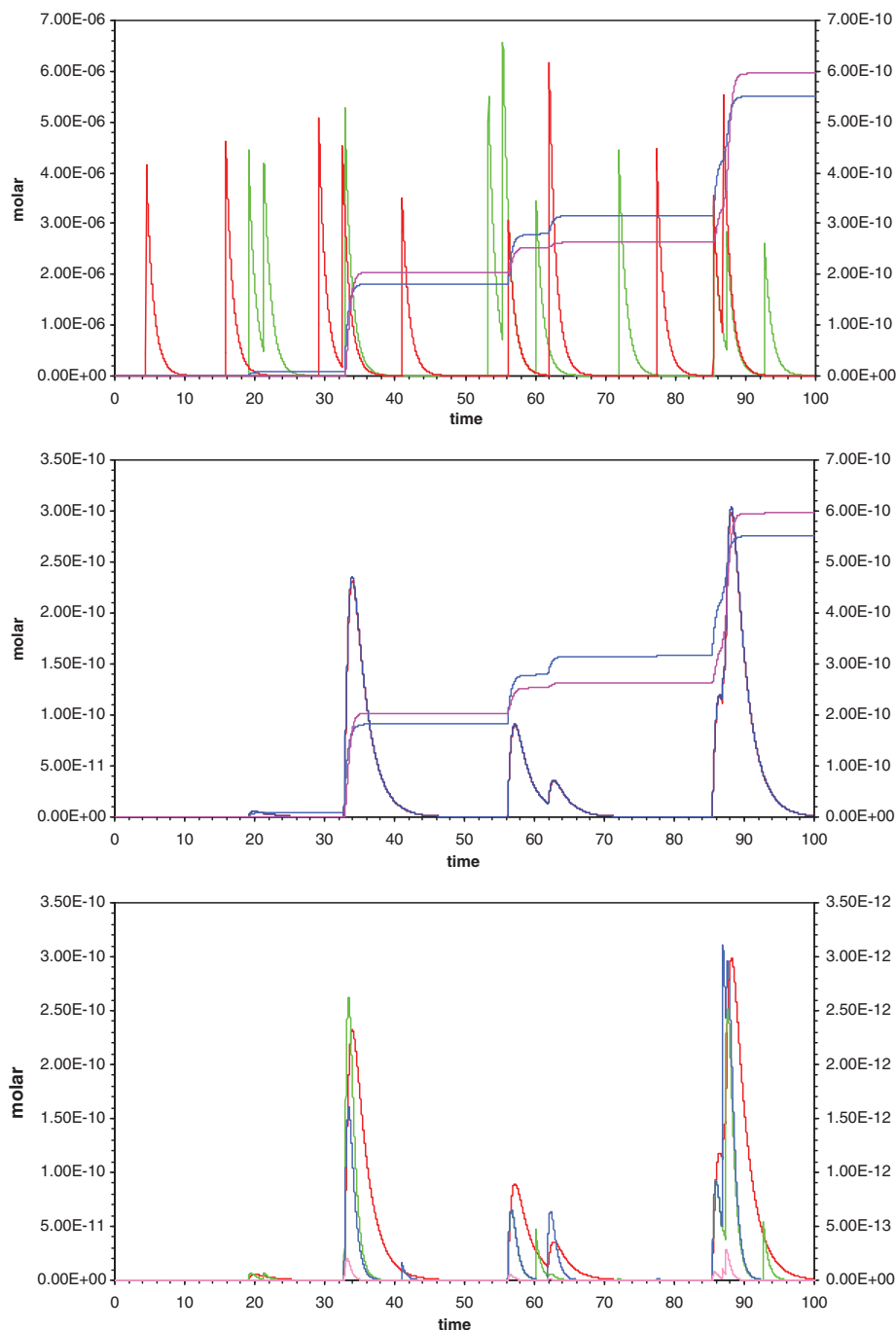


FIG. 2. One particular kinetics for all pool reactants, under standard conditions, over 100 lifetimes. (**Top panel**) *Left axis:* green and red spikes of A and B, respectively. *Right axis:* blue—total untemplated A-B synthesis; magenta—total templated A-B synthesis. (**Middle panel**) *Left axis:* blue, total instantaneous Σ A-B; red, free A-B (superposed). *Right axis:* untemplated and templated A-B, as in top panel. (**Lower panel**) *Left axis:* red, free A-B. *Right axis:* green, A-B(A); blue, (B)A-B; pink, (B)A-B(A).

data for this reaction will be in blue below), they create the first molecules of the chemically capable replicator, A-B, shown at the left. All forms of A-B are themselves unstable, decaying at $k_{\text{decay}2}$. Finally, paired dimers decay at $k_{\text{decay}3}$ to release an undamaged strand. "Decay" is meant to include hydrolysis, but also perhaps adsorption to surfaces or other first-order side reactions like cyclization.

Replication

A-B can replicate via the magenta arrow in Fig. 1 (data for this reaction in magenta below), after acting as a template which pairs B with A and thereby brings reactive precursors together in the base-paired complex (B)A-B(A) (Fig. 1). To predict the outcome, we solve the model equations suggested by Fig. 1 to determine whether A-B proliferation is likely. For Darwinian evolution, A-B must replicate, so that its information is derived from preexisting templates; therefore requirements for replication are said below to "define the Darwinian domain."

Results

Solving the system

The method of solution is simultaneous numerical integration of the 18 reactions of Fig. 1, which is detailed in **Calculations**, below and online in Numerical Results and Berkeley Madonna Simulation Code (see Supplementary Information at www.liebertonline.com/ast). A set of typical panels that follows the pool for 100 mean lifetimes for A and B is shown in the three panels of Fig. 2. Here, at a low frequency of random substrate appearance (averaging 10 reagent spikes in 100 lifetimes) are the integrated concentrations of all species in Fig. 1. Scales for reactant molar concentrations are at the right and left, for a system in which A and B appear at random, uncorrelated times at a Gaussian mean (\pm standard deviation) concentration of 4 (\pm 1) micromolar.

Pool evolution is episodic

Sporadically fed pool dynamics (Fig. 2) are dominated by the fact that reactants strongly vary in the short term but are determined when averaged over a large number of lifetimes, and secondly that all species decay at rates that embody their finite stabilities. The angled declines on the right of every peak in Fig. 2 reflect those inevitable first-order decays, as well as sporadic syntheses and dissociations. Inputs are therefore balanced with decays—the long-term solution for the pool is an "unsteady state" reflecting poise between rates of substrate appearance and universal decay.

However, the special nature of this unsteady state is of great consequence for pool chemistry. In Fig. 2, the random nature of precursor supply means that, usually, spikes of one substrate decay before a spike of the other. This ensures that both templated and untemplated A-B synthesis take place in relatively short intervals, at times when A and B overlap by chance. Instability of all species ensures that always, after episodes of synthesis, the results decay, until the next sporadic synthesis occurs. Thus, the history of the pool is sporadic A-B synthesis, at varying spacing in time, interleaved with (potentially) prolonged decays (Fig. 2, all panels).

Decay sets the timescale in the following sense: RNA bases are fairly stable—at pH 7 and 0°C, G has a mean lifetime of

1.9×10^6 years (Levy and Miller, 1998). C, the least stable base because of facile deamination to U, has a mean lifetime of 2.5×10^4 years. However, the mean life of a normal 3',5' ribonucleotide phosphodiester bond is only about 95 years at 0°C and pH 7 (calculated from data of Soukup and Breaker, 1999). Least stable of all is ribose, however, with a mean life of 63 years (Larralde *et al.*, 1995) at pH 7 and 0°C. Activated nucleotide phosphates are less stable yet, with nucleotide 5' imidazolides having a mean lifetime of the order of 10 days at lowered temperature. Thus I adopt the shortest mean lifetime of 10 days for calculation (mean lifetime = $\text{rate}^{-1} = t_{1/2} / \ln 2$). Under this assumption, Fig. 2 spans about 3 years in the life of the pool. However, other similar lifetimes could clearly be justified, expanding or contracting the time spanned by a typical simulation. Any plausible lifetime, however, puts pool behavior at the far edge of current experimentation, in a range that only calculation can thoroughly explore.

I emphasize one important numerical implication: consider a peak of A-B as a "trial" (Fig. 2, middle), that is, an opportunity for assembly of a more complex biological network utilizing the chemical capabilities of A-B. Then if one has several $\times 10^8$ years to find a successful combination, there will be of the order of hundreds of millions of trials *per currently active pool*. Thus, quick or probable access to an IDA is not a hidden assumption in these results.

Episodic synthesis during random overlaps between substrates shapes the overall dynamics of A-B appearance. We focus on the domain of replication, where the majority of A-B comes from templated synthesis (see below)—past this point most A-B information comes from previous A-B template, and (if there are differing candidates for, *e.g.*, B) Darwinian selection can accompany A-B propagation. The mean integrated appearance of A-B is not exponential in time but linear after an accelerating beginning. That is, episodes of synthesis vary in spacing but after 100 lifetimes have occurred at a determined mean rate that increases with the 3.76 power of the A and B spike frequency (see **Calculations**). After an acceleration due to initiation at zero concentrations of all reagents (and therefore a zero initial rate of synthesis), total templated and untemplated A-B rise at a varied, sporadic, but overall linear mean rate. This behavior is seen in total integrated templated A-B versus time, plotted in Fig. 3 as the mean and standard deviation of 10 standard initiations, through 100 lifetimes.

A-B is largely free

In the middle panel of Fig. 2, total A-B in all forms has been plotted with free A-B on the same scale. The two plots superpose—only one line is visible, except during the greatest transients or peaks. Similar superposition occurs under all conditions below, with the implication that most A-B is free for biochemistry or replication, not tied up in complex. While there is no reason a base-paired A-B dimer cannot be chemically active on its unpaired faces, and so also have a phenotype, the standard dinucleotide system analyzed is not dominated by template base pairing. It needs no helix-destabilizing activity to free replicated strands for use as, for example, replication templates (von Kiedrowski, 1986).

Over a broad range of conditions, the oligonucleotide pool is therefore a system that replicates even though its template is always substantially free (Fig. 2, Fig. 4). A worst case can be found at the rightmost peak of A-B in Fig. 2 (where

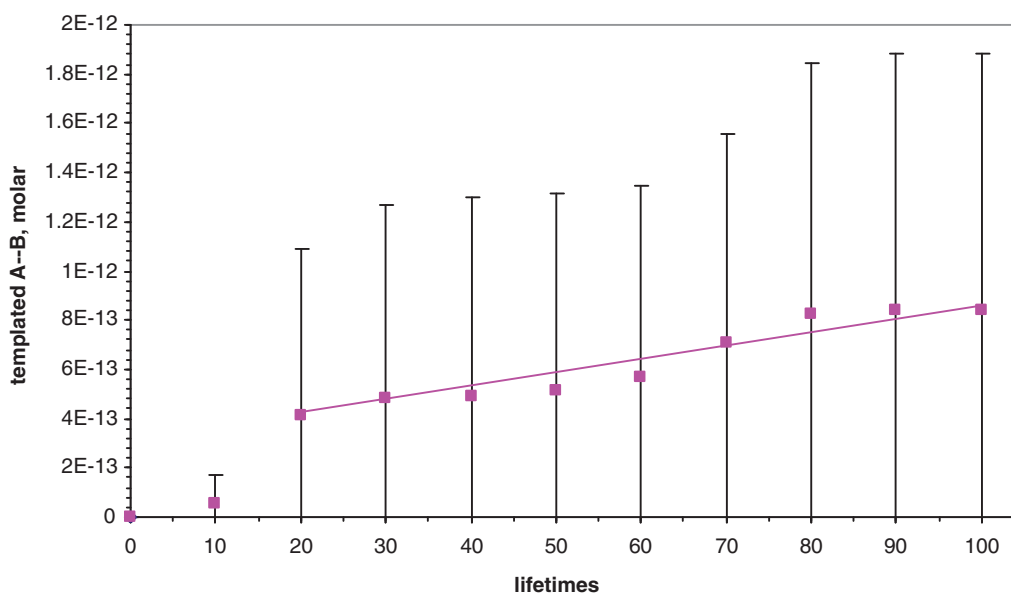


FIG. 3. Mean templated A-B, with standard deviation bars, in 10 averaged simulations, showing an accelerating start and later linearity with time.

potential ligands for A-B are near maximal); the dimer is 98.3 % free. This differs from larger nucleic acids, even from self-complementary oligonucleotides only slightly larger (von Kiedrowski, 1986; Sievers and von Kiedrowski, 1994).

Predominantly free A-B is at the root of a number of important pool properties. For example, the rate of templated A-B synthesis is first order (proportional to) total or free A-B concentration, when traced through a peak in Fig. 2. This is necessary and sufficient to ensure that A-B is a replicator whose properties are optimal for change under selection (Szathmáry and Maynard Smith, 1997). As expected, direct chemical synthesis of A-B is, in contrast, zeroth order (independent of) A-B concentration. Thus as replication increases, a new pool is first independent of preexisting A-B concentration, then approaches first-order dependence on its A-B when templating becomes the dominant mode of A-B production. Whenever variance in behavior and selection align during replication, pool properties should evolve rapidly. Because replicator evolution can sometimes be related to differential molecular stabilities (Lifson and Lifson, 1999; Scheuring and Szathmáry, 2001), standard pool A-B was also checked through the point of replication with the same lifetimes for all reactants. Free A-B was observed throughout. Not only is free A-B crucial to selection, below we also find that its freedom determines dynamic pool behavior.

Peak A-B

The other quantity presented in detail (*e.g.*, in Fig. 4) is peak A-B; the maximum A-B present during the most effective episode of synthesis. Peak A-B (*e.g.*, the rightmost peak in Fig. 2, middle and bottom) is the maximum molecular concentration of A-B in all forms, with all kinetic factors taken into account. By contrast, untemplated A-B is the integrated amount made by direct reaction between A and B (with no decay), a quantity useful for discussion of the effectiveness of untemplated synthesis. Templated A-B is,

similarly, the integrated amount of A-B made on a pre-existing A-B template (with no decay), a quantity useful for discussion of the effectiveness of replication. Put another way, templated and untemplated A-B are the total amounts that would be present if all decay were negligible.

Figure 2 shows that total A-B synthesis in 100 lifetimes occurs in a small number of episodes. An episode's contribution can be close to the amount plotted as peak A-B (Fig. 4). Thus the black line (peak A-B) is usually within a factor of 2 or 3 of totaled synthesis [total untemplated A-B (blue) plus templated A-B (magenta)]. Almost without exception here, total A-B synthesis during 100 lifetimes has occurred in one, two, or three episodes, though an episode can be complex, entraining several substrate spikes, as seen here for the latter two episodes in Fig. 2.

Intrinsic stability selection in the pool

Episodic synthesis has another important implication. Deep troughs appear between episodic peaks, where many orders of magnitude in template concentration are lost (Fig. 2, all panels). Therefore, there is unavoidably rigorous selection for A-B stability. If a sporadically fed pool contains A and B variable in stability (and stability is expressed in the oligomer), more stable variants will rapidly be selected. As we will see below (Fig. 4C), this rapidly elevates production of A-B and increases replication, bringing on transition to the Darwinian domain. Afterward, synthetic episodes usually begin with preexisting templates selected for stability. The basis for this tendency is already visible in episodes that begin just past 56 and 85 lifetimes in Fig. 2; in each case A-B is continuously present and actively serving as template for ≈ 15 lifetimes, through several spikes. Further examination of such a transition to increased stability seems likely to provide further useful information.

Possible fossilization

Figures 2 and 4 implicitly offer a possibility for experimental detection of an ancient sporadically fed pool where

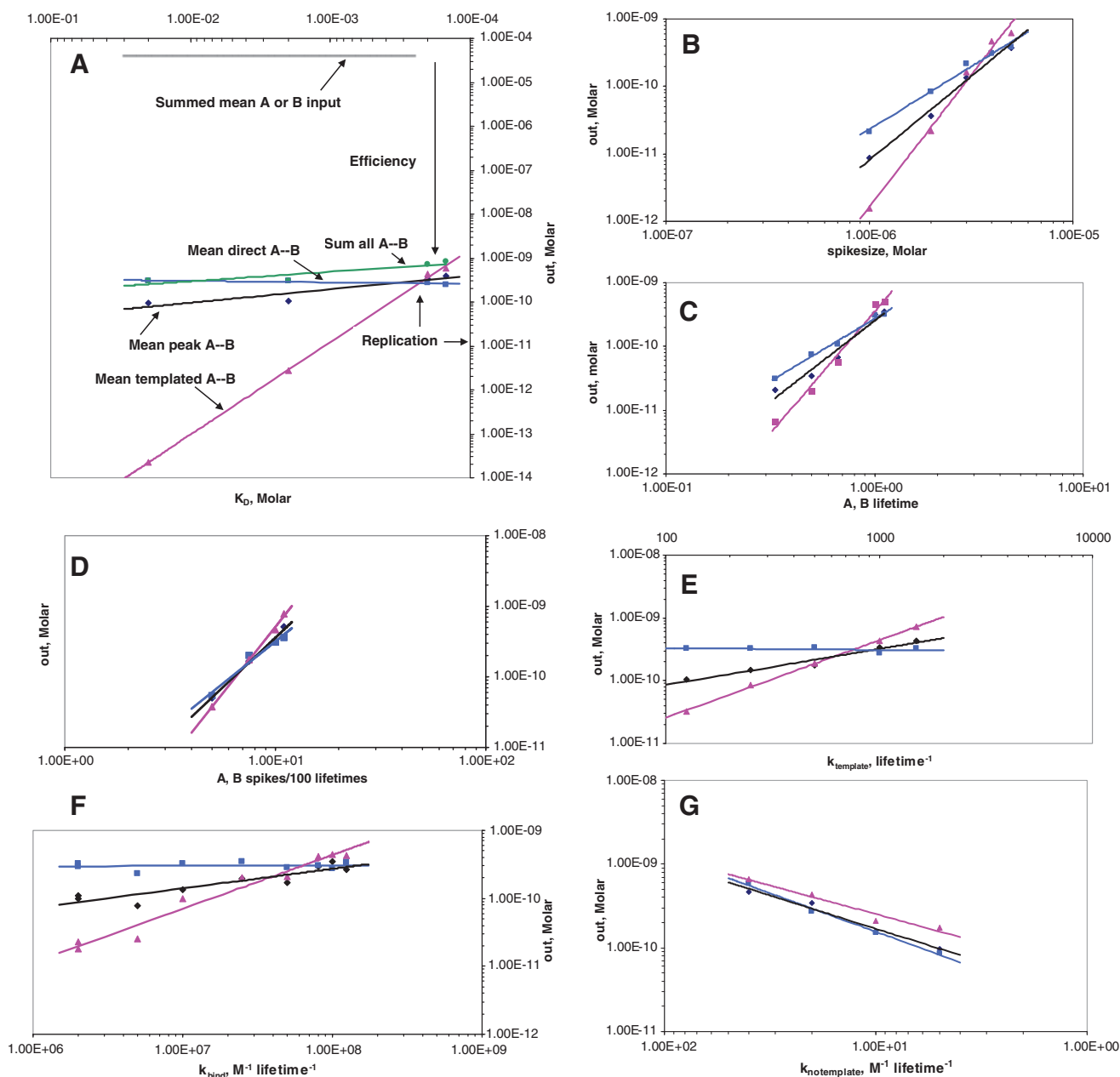


FIG. 4. (A) Effect of the nucleotide dissociation constant (K_D) on mean direct A--B (blue), mean templated A--B (magenta) and mean summed A--B (untemplated + templated, green), as well as mean peak A--B concentration (black) after 100 lifetimes; average of 25 simulations. Effect of substrate **spikesize** (B) and molecular **lifetime** (C) on mean direct A--B synthesis (blue), mean peak A--B synthesis (black), and templated A--B synthesis (magenta), after 100 lifetimes; averages of 25 simulations. (D) Effect of mean substrate spike frequencies on mean direct (blue), mean templated (magenta), and mean peak A--B (black) synthesis from the pool after 100 lifetimes. (E) Effects of the rate of templated dimer formation (k_{template}) on mean direct synthesis (blue), mean templated synthesis (magenta), and mean peak A--B synthesis (black) after 100 lifetimes; averages of 25 simulations. (F) Effects of the rate of nucleotide association (k_{bind}) on mean direct (blue), mean templated (magenta), and mean peak A--B synthesis after 100 lifetimes; averages of 25 simulations. (G) Effects of the rate of untemplated AB synthesis ($k_{\text{notemplate}}$) on mean direct (blue), mean templated (magenta), and mean peak A--B after 100 lifetimes; averages of 25 simulations.

Darwinian life might have originated. Successive layered episodes along the time axis might fossilize, if the pool communicates with a sediment-forming zone. Synthetic episodes necessarily contain simultaneous peaks of both substrates and products (Fig. 2), so could be recorded as strongly varying concentrations of elemental phosphate in sedimentary rocks and perhaps detected after an indefinite time. It is intriguing to think that sporadic pool kinetics and (e.g.) mi-

croprobe mass spectrometry might define possible regions for life's origin and even tell us something about its kinetics, despite intervening gigayears.

A quantum for creation

A final aspect of pool A--B synthesis discernible from overall inspection of Figs. 2 and 4 is that there is a minimum

episode of synthesis, which is not decreased by, for example, further decreases in substrate frequency. That is, overlap between substrates becomes less frequent as spike frequencies decrease (trend shown explicitly in Fig. 4C), but both the average and maximum size of an A-B creation episode have a lower limit, determined by the integrated overlap between spikes. The average of 250 completely isolated, *simultaneous* A and B spikes generated here had a yield of 0.55 ± 0.04 nM A-B (mean \pm standard error of the mean). With less frequent supplies of A and B, pool tempo alters in that overlaps become less frequent, but neither the distribution nor peak yields from an episode will decrease. The maximal synthetic episode, from precisely overlapping spikes, will always occur if one has time to wait.

Calculations

Synthesis and decay of a self-complementary A-B dimer in a sporadically fed pool (Fig. 1) were embodied in the algebraic and differential equations in Fig. 5. This system of differential equations was integrated and plotted by using Rosenbrock integration (Rosenbrock, 1963) at 10^3 intervals/lifetime, employing Berkeley Madonna v 8.3.18 (R.I. Macey

and G.F. Oster, University of California) running on a Lenovo T61 computer (simulation code and results are in Supplementary Information, available online at www.lieberonline.com/ast). Further analysis of simulation data utilized Microsoft Excel 2003, SP3. Runs of 100 lifetimes begin with all reagent concentrations set to zero, and subsequent randomly-timed independent Gaussian-distributed spikes of A and B (**RIG**) appear in the pool with a chosen mean frequency (in the standard example a mean peak of $4 \mu\text{M}$ with a standard deviation of $1 \mu\text{M}$ at a mean of 0.1 spikes per lifetime). Figure 4 points are means of 25 such simulations for 100 lifetimes.

Standard errors for total direct A-B synthesis over 100 lifetimes are well determined, with standard errors frequently $\approx 10\%$ of the mean, but peak and templated A-B have large variance, often of the order mean². Note the standard deviations in Fig. 3. However, the averaging process adopted gives relatively reliable behavior (data in Fig. 4 panels). Other integration methods and variations of the integration interval give similar numerical results; thus integrations appear quantitatively dependable.

The last three equations of Fig. 5 define quantities given particular emphasis in discussion. Proceeding downward, these three equations describe the total synthesis of oligomer

$$\begin{aligned} \frac{dA}{dt} &= -k_{bind} \cdot [A--B \cdot A + (B)A--B \cdot A] - k_{decay1} \cdot A - k_{notemplate} \cdot A \cdot B + k_{diss1} \cdot A--B(A) + k_{diss2} \cdot (B)A--B(A) + \mathbf{RIG} \\ \frac{dB}{dt} &= -k_{bind} \cdot [A--B \cdot B + A--B(A) \cdot B] - k_{decay1} \cdot B - k_{notemplate} \cdot A \cdot B + k_{diss1} \cdot (B)A--B + k_{diss2} \cdot (B)A--B(A) + \mathbf{RIG} \\ \frac{dA--B}{dt} &= -k_{bind} \cdot [A--B \cdot A + A--B \cdot B] - k_{decay2} \cdot A--B + k_{notemplate} \cdot A \cdot B + k_{diss1} \cdot [(B)A--B + A--B(A)] \\ &\quad + k_{decay3} \cdot (A--B)(B--A) + 2 \cdot k_{diss3} \cdot (A--B)(B--A) - k_{bind} \cdot (A--B)^2 \\ \frac{d(B)A--B(A)}{dt} &= -2 \cdot k_{diss2} \cdot (B)A--B(A) - k_{decay2} \cdot (B)A--B(A) - k_{template} \cdot (B)A--B(A) + k_{bind} \cdot (B)A--B \cdot A \\ &\quad + k_{bind} \cdot A--B(A) \cdot B \\ \frac{dA--B(A)}{dt} &= -k_{bind} \cdot A--B(A) \cdot B - k_{diss1} \cdot A--B(A) - k_{decay2} \cdot A--B(A) + k_{bind} \cdot A--B \cdot A + k_{diss2} \cdot (B)A--B(A) \\ \frac{d(B)A--B}{dt} &= -k_{bind} \cdot (B)A--B \cdot A - k_{diss1} \cdot (B)A--B - k_{decay2} \cdot (B)A--B + k_{bind} \cdot A--B \cdot B + k_{diss2} \cdot (B)A--B(A) \\ \frac{d(A--B)(B--A)}{dt} &= k_{template} \cdot (B)A--B(A) + k_{bind} \cdot (A--B)^2 - k_{decay3} \cdot (A--B)(B--A) - k_{diss3} \cdot (A--B)(B--A) \\ \left. \frac{dA--B}{dt} \right]_{\text{untemplated}} &= k_{notemplate} \cdot A \cdot B \\ \left. \frac{dA--B}{dt} \right]_{\text{templated}} &= k_{template} \cdot (B)A--B(A) \\ \text{total} \cdot \text{template} &= A--B + A--B(A) + (B)A--B + (B)A--B(A) + 2 \cdot (A--B)(B--A) \end{aligned}$$

FIG. 5. The pool mechanism of Fig. 1 formulated as differential equations, to be solved by numerical integration. **RIG** indicates Random Independent Gaussian spikes of A and B. Dots indicate algebraic multiplication, and parentheses [for example, (B)A-B] indicate noncovalent, base-paired ribonucleotide complexes. Dimer notation, A-B, memorializes the 5'-5' cofactor molecules whose potential possession of both catalytic and replication properties (Yarus, 2011a) inspired this development; but A and B could represent other modules, perhaps multinucleotide, perhaps having 5'-3' backbones. However, A and B are always complementary; (A-B)(B-A) is the base-paired self-complementary dimer. Note that for cofactor-like molecules, A-B \equiv B-A.

by direct reaction between nucleotides [used to define (blue) untemplated A-B synthesis], then the total synthesis of oligomer by templated reaction [used to define (magenta) effects on A-B replication], and last, the instantaneous summed concentration of all oligomer (black; totaled A-B peaks exemplify the maximum contribution of the pool to its environment).

Detailed results and individual variations

In the panels of Fig. 4, log-log plotted lines are least-squares lines fit to data points (themselves the means of twenty-five 100 lifetime simulations; see Numerical Results, Supplementary Information online).

The first panel (Fig. 4A) has additional labeling to clarify interpretation. At the top, the standard total peak input per 100 lifetimes (a mean of 10 spikes with a mean peak size of 4 μM) is shown. In most panels this line is flat at $4 \times 10^{-5} M$, as in Fig. 4A, but when peak substrate is varied, input has a slope of unity (linearly increasing; Fig. 4B and 4D).

In Fig. 4A, the ratio of templated, direct, or templated plus direct A-B synthesis to total input can be thought of as an efficiency, or yield, of A-B. Thus we say of panel 4A that the efficiency of total A-B synthesis grows as nucleotide binding increases. In fact, this is important in discussion because it is true in Fig. 4A through Fig. 4F—that which increases replication also, in almost all cases, increases the total yield of A-B dimer (the exception being $k_{\text{notemplate}}$; Fig. 4G).

Where the blue (direct reaction of A and B) and magenta (templated reaction of A and B) lines cross is the point at which integrated untemplated synthesis and integrated templated synthesis are equal. Thus in Fig. 4A decreasing K_D is favorable to replication (templated A-B has the greater slope), and to the right of the crossover (arrows marked “replication” in the panel), replication is the dominant mode of A-B synthesis. Thus, a blue-magenta pencil of lines that opens to the right means that replication is favored by decreasing K_D ; and below $K_D = 2.3 \times 10^{-4} M$ (under standard conditions; see Table 1 and Calculations, below), we are in the domain of replicating, potentially Darwinian, efficiently selected systems. Other panels are plotted similarly; replication increases rightward, and the Darwinian transition occurs midpanel.

Finally in Fig. 4A, mean peak dimer (black line) parallels total synthesis (green line) across all tested spike magnitudes. Because the largest synthetic peak is a large, near-constant fraction of total A-B synthesis, the episodic nature of synthesis, which occurs in only a few events in 100 lifetimes, is preserved across the range explored. To increase clarity, total synthesis data (green) are not plotted in other panels, but

similar episodic synthesis (almost parallel, close, total A-B and peak A-B lines) is the rule throughout.

Individual pool quantities

We now consider the sporadically fed pool, quantity by quantity Q (e.g., spikesize, k_{template} ...). Reaction variables are considered in order of importance to replication, with most influential first. I use least-squares lines for simulation data (Fig. 4 and Calculations) to represent pool behavior as quantity Q varies:

$$A - - B_{\text{Untemplated}} = c_{\text{Untemplated}} Q^{Y_{\text{Untemplated}}}$$

$$A - - B_{\text{Templated}} = c_{\text{Templated}} Q^{Y_{\text{Templated}}}$$

In particular, the variation of replication as pool quantity Q is varied is

$$\begin{aligned} A - - B_{\text{Templated}} / A - - B_{\text{Untemplated}} \\ = [c_{\text{Templated}} / c_{\text{Untemplated}}] Q^{Y_{\text{Templated}} - Y_{\text{Untemplated}}} \end{aligned}$$

And the exponent $Y_{\text{Templated}} - Y_{\text{Untemplated}}$ can be used as an index of Q 's influence on replication after 100 lifetimes. Greater exponents mean sharper dependence and more influence. Q_{Darwin} , where total untemplated and templated A-B synthesis are equal is

$$Q_{\text{Darwin}} = [c_{\text{Untemplated}} / c_{\text{Templated}}]^{1/(Y_{\text{Templated}} - Y_{\text{Untemplated}})}$$

Past Q_{Darwin} (rightward in all panels of Fig. 4), replication of A-B is predominant over untemplated synthesis, and I write of being in the Darwinian domain.

Nucleotide dissociation constants

K_D is the most significant quantity to replication. Fig. 4A's abscissa decreases to the right (better nucleotide binding to A-B), so that replication increases rightward, as in other panels. Replication increases as $K_D^{-2.15}$, making the nucleotide dissociation constant the most influential pool variable.

To allow k_{bind} (the rate of base pairing) to vary without unrealistic variation in the equilibrium binding constants of nucleotides, dissociation rates for nucleotides bound to A-B, k_{diss1} (for the first nucleotide to bind) and k_{diss2} (for the second nucleotide to bind to a stack initiated by the first nucleotide) are changed concurrently with k_{bind} so that the equilibrium dissociation constant for the first nucleotide is always $K_1 = k_{\text{diss1}} / k_{\text{bind}} = 0.2 \times 10^{-3} M$, and the dissociation rate for the second nucleotide is maintained so that $K_2 = k_{\text{diss2}} / k_{\text{bind}} = 0.2$

TABLE 1. STANDARD VALUES USED IN CALCULATIONS OF THE PROPERTIES OF A SPORADICALLY FED POOL, AND CALCULATED LIMITS FOR POOL REPLICATION (HEADED “REPLICATION”), THE LATTER (Q_{DARWIN}) IN TIME UNITS OF BOTH SUBSTRATE A AND B LIFETIMES AND SECONDS

Pool Q varied	Standard value	Replication (Q_{Darwin})	Replication (t=seconds)	Comment
Mean spike	$4 \times 10^{-6} M$	$\geq 3.6 \times 10^{-6} M$	—	Mean A = mean B
Mean lifetime	1/0.5/0.25	$\geq 0.87/0.44/0.22$	—	Locked together
Mean spike freq	$0.1 \text{ lifetime}^{-1}$	$\geq 0.071 \text{ lifetime}^{-1}$	$\geq 8.2 \times 10^{-8} \text{ s}^{-1}$	Mean A = mean B
K_D	$0.2 \times 10^{-3} / 10^{-4} / 10^{-5}$	$\leq 0.23 \times 10^{-3} / 10^{-4} / 10^{-5}$	—	Locked together
k_{template}	$1000 \text{ lifetime}^{-1}$	$\geq 750 \text{ lifetime}^{-1}$	$\geq 8.7 \times 10^{-4} \text{ s}^{-1}$	
$k_{\text{notemplate}}$	$10 M^{-1} \text{ lifetime}^{-1}$	$\leq 79 M^{-1} \text{ lifetime}^{-1}$	$\leq 9.1 \times 10^{-5} M^{-1} \text{ s}^{-1}$	
k_{bind}	$1 \times 10^8 M^{-1} \text{ lifetime}^{-1}$	$\geq 6.4 \times 10^7 M^{-1} \text{ lifetime}^{-1}$	$\geq 74 M^{-1} \text{ s}^{-1}$	

$\times 10^{-4}$ M. The self-association of A-B is similarly locked to k_{bind} so that the dissociation constant for self-complementary dimer is always 0.2×10^{-5} M. Under these conditions, the Darwinian domain is $K_D \leq 2.3 \times 10^{-4}$ M for single nucleotides, aided slightly because direct synthesis of A-B *decreases* slightly as K_D falls and A-B template becomes increasingly able to capture free A and B for templated synthesis (Fig. 4A).

These K_D are chosen to give weak but arguably realistic binding and to reflect well-known stacking and entropic phenomena of nucleotide base-pair chemistry. For example, they anticipate better binding when stacking the second nucleotide on the first-to-bind, as well as tighter binding of the covalent dimer to itself in 3'-5' backbones (Freier *et al.*, 1986). In addition, the factor of 10 between isolated nucleotide pairing to template and second, adjacent nucleotide binding adopts the factor deduced from kinetics of ImpG reaction on poly C (Kanavarioti *et al.*, 1999). However, such stabilities are not well defined, particularly for 5'-5' RNA backbones, a focus here. An interesting possibility, suggested by the selective stimulation of particular base pairings by adjacent modified nucleotides in the anticodon loops of tRNAs (Agris *et al.*, 2007), is that the chemically active B nucleotide's modification(s) might also stimulate base pairing (lower K_D) during templated synthesis. In any case, definite binding for single nucleotides is implied. Thus, there is a clear need for definitive biochemistry on this, the most influential pool quality. Variation of K_D is discussed further in *Effect of available spikesize*, below.

Mean spike size

The size of substrate spikes is the second most significant quantity for replication. Substrate supply in Fig. 4B also deserves discussion because bigger spikes are an influential way to elevate the A-B yield and degree of pool replication. Templated A-B output increases as **spikesize**^{3.9}, evocative of the four aliquots of substrate used to replicate. Replication increases as **spikesize**^{2.05}, second from the maximum dependence observed. Thus, Darwinian behavior and efficient selection can begin at low pool nucleotide precursor concentrations, low even by present biological standards. However, pool output above the Darwinian boundary at **spikesize** ≥ 3.6 μM will likely reward further exploration because higher output is related to the intensity of the pool phenotype, keeping in mind that A-B is a catalyst, and turns over.

These two most influential pool quantities, K_D and **spikesize**, directly determine the pairing of A and B to template. Template pairing is unique to replication, thus pairing is a crucial event in the Darwinian fate of the pool. Both K_D and **spikesize** also influence replication as near-powers-of-two, reflecting two successive template pairing events that together create a kinetically competent template. Because A-B dimer is mostly free, binding is weak, and therefore increased substrate increases individual binding events linearly. As a result, replicating template increases nearly as the square of effective substrate supply, and the underlying stoichiometry shines through, despite numerous mechanistic and kinetic complexities in the life of a pool (Figs. 1 and 2).

Molecule lifetimes

Both templated (\propto **lifetime**^{3.82}) and direct A-B syntheses increase strongly when the coupled reagent lifetimes increase,

and replication increases also, rising as **lifetime**^{1.85} (Fig. 4C). For simplicity, A and B are assumed to have the same 10-day lifetime, though A contains, by hypothesis, an activated phosphate while B is a possibly more stable unactivated dinucleotide. The domain of replication for the standard case begins slightly below standard conditions (**lifetime** = 1), when A and B survive 0.87 lifetimes (compare **Calculations**). In the model, decay rates are coupled; all forms of the A-B dimer, paired and unpaired to monomer(s), are given the same lifetime, twice that of free reactants A and B. Slightly longer life for A-B is consistent with the usual observed stabilization of nucleic acids by increased structure (Lindahl, 1993). For the same reason, the lifetime for base-paired dimer (A-B)(B-A) is maintained at four times that of A and B, or twice that of dinucleotide A-B. Effects of increased stability are mostly mediated by the stability of the substrates A and B, rather than by forms of A-B. However, because A-B is no longer activated, a model with much longer lifetimes for its congeners than assumed here would also be justifiable.

Long molecular lifetime has particular importance for the history of a pool; if different forms of A and B with varied stabilities coexist, such an initially mixed pool can rapidly be selected to climb rightward in Fig. 4C, toward increased yield and replication. Evolution of more stable A-B is also of particular interest because selected A-B potentially records earlier selection of all types and potentially allows pool properties to evolve coherently through extended times. Conditions for transition to coherent long-term evolution seem worth further exploration.

Mean spike frequency

The next most influential replication variable is the frequency of substrate appearance. Total templated pool output goes as **freq**^{3.76}, as might be expected from the importance of spikes to both the undirected synthesis of template and in its subsequent replication. However, because untemplated reaction to form dimer A-B is also highly frequency-dependent, increasing **freq** is less effective in promoting replication than is spike size; replication increases as **freq**^{1.37}. Replication begins at 0.071 spikes/lifetime on average (or 8.2×10^{-8} spikes s^{-1}), which is quite infrequent (see the blue-magenta crossover in Fig. 4D). A and B substrates therefore decay to 7.6×10^{-7} of their peak concentrations, on average, between their successive peaks, but replication nevertheless persists.

Templated reaction rate

The first-order rate, k_{template} , controls formation of A-B after base-paired nucleotides are aligned on an A-B template. Overall templated synthesis of A-B is moderately dependent on the rate of this bond formation, with total synthesis of templated A-B increasing as k_{template} ^{1.23} and replication increasing as k_{template} ^{1.27}. Increasing k_{template} (Fig. 4E), as also seen in Fig. 4D, slightly decreases untemplated synthesis by capturing free A and B nucleotides for templated synthesis.

In Fig. 4E, k_{template} varies from 125 to 1500 lifetime^{-1} . At 10 days/lifetime, this range is 1.4×10^{-4} to 1.7×10^{-3} s^{-1} and is 1.2×10^{-3} s^{-1} when not being specifically varied. The poly U-templated reaction of AMP carbodiimide with A occurs at greater than 6×10^{-5} s^{-1} , based on the data of Sulston *et al.* (1968), though this may take place in a triple U₂A helix. More

closely related, therefore, is the formation of poly C-templated 5'-3' G dimers using 2-MeImpG at $5 \times 10^{-5} \text{ s}^{-1}$ (Kanavarioti *et al.*, 1999) as well as the related 5'-3' addition of paired 2-MeImpG to an oligo-G primer on a poly C template, at $6.9 \times 10^{-4} \text{ s}^{-1}$ (Kanavarioti *et al.*, 1998). Accordingly, present calculations overlap rates accessible to imidazolidine-activated G within 2'-5' and 3'-5' backbones, though the unknown rates for templated 5'-5' reactions would be even more germane. Under standard pool conditions, the Darwinian domain lies ≥ 750 per lifetime, or $\geq 8.7 \times 10^{-4} \text{ s}^{-1}$ at a lifetime of 10 days.

It is interesting to compare triphosphate-mediated ribonucleotide ligation with these calculations. Rohatgi *et al.* (1996) studied a potentially comparable templated, but minimally catalyzed, ribonucleotide 5'-3' ligation reaction. Fastest observed ligation rates were $6.9 \times 10^{-8} \text{ s}^{-1}$, in mixtures of Mg^{2+} and Pb^{2+} at pH 7.4 and 37°C. Thus, minimally catalyzed triphosphate-mediated 5'-3' joining is more than four orders below the Darwinian boundary of the sporadically fed pool. This suggests the need for a more reactive primordial phosphate-activating group than pyrophosphate. Such a requirement is not an accident but a consequence of the stability of phosphate diesters in the absence of specific catalysis (Westheimer, 1987). Rates required for replication are instead close to those measured for imidazolidine-activated 5'-3' nucleotide joining.

Nucleotide pairing

In Fig. 4F, the effects of k_{bind} , the second-order rate at which all nucleotides pair, are shown. In these calculations, k_{bind} varies from 2×10^6 to $1.1 \times 10^8 \text{ M}^{-1} \text{ lifetime}^{-1}$, or 2.3 to $127 \text{ M}^{-1} \text{ s}^{-1}$ (for a lifetime of 10 days). These values are much slower than experimental rate constants of $\approx 3 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ (Labuda *et al.*, 1985).

Variation in this rate hardly alters untemplated synthesis at all, as would be expected because this pathway uses no pairing. Template activity instead increases with faster pairing, but slowly—mean peak A-B increases as $k_{\text{bind}}^{0.29}$ and mean templated A-B (replication) as $k_{\text{bind}}^{0.79}$. Replication in the standard pool increases at $k_{\text{bind}}^{0.784}$ and requires $k_{\text{bind}} \geq 6.4 \times 10^7 \text{ M}^{-1} \text{ lifetime}^{-1}$, or $\geq 74 \text{ M}^{-1} \text{ s}^{-1}$. This is several orders slower than experimental rates for base pairing with complementary nucleotides (Labuda *et al.*, 1985), so access to replication and Darwinian behavior via nucleotide pairing seems assured.

Untemplated reaction rate

This is the untemplated, second-order, solution reaction of A and B to give A-B. In Fig. 4G, $k_{\text{notemplate}}$ is varied from 5 to $40 \text{ M}^{-1} \text{ lifetime}^{-1}$. Taking a "lifetime" as 10 days, this range is 5.8×10^{-6} to $4.6 \times 10^{-5} \text{ M}^{-1} \text{ s}^{-1}$. When not specifically varied, $k_{\text{notemplate}}$ is set to $20 \text{ M}^{-1} \text{ lifetime}^{-1}$, or $2.3 \times 10^{-5} \text{ M}^{-1} \text{ s}^{-1}$.

These rates resemble those for phosphorimidazolides. For example, the untemplated rate at which imidazolyl-A reacts to form oligomeric ribonucleotides is $\approx 2 \times 10^{-5} \text{ M}^{-1} \text{ s}^{-1}$ in Mg^{2+} and more than 10-fold faster in 25 mM Zn^{2+} , as estimated from the data of Sawai and Orgel (1975). More explicit data from von Kiedrowski (1986) agree, putting the rate of untemplated imidazolyl-oligonucleotide reaction to form 2'-5' and 3'-5' backbones at $3 \times 10^{-5} \text{ M}^{-1} \text{ s}^{-1}$.

Unusually among pool Q, $k_{\text{notemplate}}$ for a 5'-5' untemplated reaction ($\text{ImpN} + \text{pN} \rightarrow \text{NppN} + \text{Im}$) has been

measured at pH 7 and 37°C (Kanavarioti *et al.*, 1992) at $4.7 \times 10^{-5} \text{ M}^{-1} \text{ s}^{-1}$ (or $41 \text{ M}^{-1} \text{ lifetime}^{-1}$). Thus the rates used in simulation are achievable by using ribonucleotide phosphorimidazolides. Even more interestingly, separate reactions yielding 2'-5' and 3'-5' backbones and 5'-5' backbones have similar rates.

The Fig. 4G abscissa declines to the right, so that replication increases rightward as in other panels. Slowing untemplated dimer formation decreases all pool outputs but forces more dependence on templated synthesis, so replication increases rightward. All pool output dependences on $k_{\text{notemplate}}$ are moderate, with exponents of 0.7–0.9. Replication is the dominant mode of A-B synthesis at less than $79 \text{ M}^{-1} \text{ lifetime}^{-1}$, or $\leq 9.1 \times 10^{-5} \text{ M}^{-1} \text{ s}^{-1}$. Replication varies as $k_{\text{notemplate}}^{-0.234}$ in the standard pool, reflecting inhibition of templated synthesis by facile consumption of nucleotides in untemplated A-B synthesis.

Mean simulation results are summarized in part in Table 1 above, which collects standard pool quantities, as well as the boundary values (Q_{Darwin}) for replication under standard conditions. However, no sooner is an explicit solution for one set of conditions seen than it evokes the need for a general solution, valid at any chosen Q. I close **Results**, therefore, with two comments about sporadically fed pools with other sets of pool Q.

Effect of available spikesize on required K_D

Consider the question: precursor spikes here are dilute, several micromolar; what could happen if a pool had access to more abundant precursors? After all, pool synthesis of A-B becomes more efficient at higher substrate concentrations (described in *Detailed results and individual variations*, above). Simple extrapolation of synthesis to much larger spikes can be inaccurate, because other reactions become limiting. For example, no special activity is required to dissociate (A-B)(B-A) in Fig. 4, because it spontaneously dissociates. However, at high-enough substrate concentrations, self-complementary template will become significantly paired (von Kiedrowski, 1986; Sievers and von Kiedrowski, 1994), reducing templated A-B output. Unfortunately, precise calculation of this effect is not possible because the present integration, though designed for stiff reaction networks like these (Rosenbrock, 1963), becomes unstable with larger spikes.

But suppose we nevertheless wish to follow up on the hint above that, to preserve replication with weaker nucleotide binding to template (higher K_D ; Fig. 4A), we could exploit the nearly equally effective increased replication associated with an increased substrate spikesize (Fig. 4B). We cannot use the fitted Q equations for K_D and spikesize to solve the simultaneous variation of two variables, because each equation depends on the Q in the other. However, re-simulation can be used: if one increases spikesize from $1 \times 10^{-6} \text{ M}$ to $5 \times 10^{-6} \text{ M}$ ($K_D = 2.3 \times 10^{-4} \text{ M}$), re-simulation shows that 5-fold increased mean substrate peak is compensated by a 3.5-fold increase in K_D , while maintaining the same replication; templated synthesis/untemplated = 1.8. Thus, substrate supply and K_D are mutually related. Near standard Q_{Darwin} , higher mean substrate spikes relax the need for low K_D (better nucleotide binding) by 70% the substrate increase. By extending such reasoning, we can progressively constrain a pool's implied substrate supply as our knowledge of pool molecules improves.

Lifetime, spike frequency, and other variations

The list of Q above offers broader clues about molecular abundance in the pool. The list is ordered by decreasing influence on replication near Q_{Darwin} and thus offers guidance about which pool quantities most readily respond to selection for abundance. Alternatively, because selection for molecular abundance is inevitable at some stage, the list can be used to discuss the robustness of replication itself.

I argue (in *Intrinsic stability selection in the pool*, above) that strong selection for longer nucleotide lifetime in the pool is inevitable. Because molecular lifetimes strongly influence replication (Fig. 4C), greater A-B stability and thus greater output, along with transition to the Darwinian domain, will be selected if relevant variance in lifetime exists. This argument can be seen to be a special case of a much more general conclusion.

Because A-B is mostly free, there is a reliable relation between the exponents for pool quantities Q for untemplated and templated synthesis (see **Supplementary: Numerical results** online). Templated A-B synthesis always increases substantially faster with pool Q than untemplated (least-squares equations have a larger exponent). This implies a sharp transition to replication and Darwinian behavior, as seen where the blue and magenta lines cross in the panels of Fig. 4. There is inevitable selection for more dimer to amplify any favorable biochemical effect of A-B. As can be appreciated in Fig. 4, selection for more total dimer (untemplated + templated) will select for replication if there is natural variance available in virtually *any* pool Q . The sole exception is the untemplated rate of A and B reaction—selection for replication *decreases* $k_{\text{notemplate}}$ (Fig. 4G) and total dimer. But what is pointed out above for K_D in Fig. 4A (in *Detailed results and individual variations*), is also true of **spikesize** in Fig. 4B, **lifetime** in 4C, **spike freq** in 4D, k_{template} in Fig. 4E, and k_{bind} in Fig. 4F—selection for total A-B (that is, for any increased biochemical effect) moves the pool rightward toward increased replication and increased Darwinian behavior.

Thus, preferred molecules in the long run are readily selected replicators, and these should arise easily because they can appear as natural chemical variants in *any* pool quantity Q (except $k_{\text{notemplate}}$). Pool molecules feel selection (see *A-B is largely free* above), so replication and the Darwinian transition should be a robust outcome, accessible via many molecular routes in a sporadically fed ribonucleotide pool.

Discussion

The method of this investigation is to use chemical kinetics and measured literature reactions to demonstrate realizable examples of system behavior. For example, it is shown that pools of small ribonucleotides plausibly exist that replicate even though its template is mostly unassociated: always free.

Other implications of these calculations

Naming the IDA is worthwhile in that it focuses thought to a defined, crucial point; the simplest or earliest system that might both replicate and express a chemical phenotype. A third plausible requirement for the IDA is continuity with the later, but arguably overlapping, RNA world (Yarus, 2011a). Notably, small RNAs exist which appear to meet the three

requirements. These are the ribodinucleotide cofactors, chemical relatives of modern NAD or FAD (Yarus, 2011a). This prior argument is carried further here, where the kinetic requirements for ribodinucleotide replication in a sporadically fed pool of unstable nucleotides are calculated. Calculations focus around the point where pool replication begins to transmit information to descendants (Q_{Darwin}) in order to discuss this event.

The crux of the argument is the numerical solution of the differential equations that constrain the sporadically fed pool (Figs. 1 and 5). To interpret the results (*e.g.*, Fig. 2), consider that the problem is not framed in a way favorable to replication. For example, substrates appear at random times, completely uncorrelated. Amounts are varied, distributed as a Gaussian and always very dilute. Substrate spikes are infrequent, typically about one in 10 lifetimes. All reactants decay in a time measured in days. Some rates, such as k_{bind} , whose increase favors replication (Fig. 4F), are taken to be orders slower than experimental values. The pool itself is mortal, disappearing when it has survived 100 times as long as its least stable component, at a time when output is still affected by an assumed start at zero rates (Fig. 3). Nevertheless, pool components begin to replicate (Table 1, Fig. 4), and this pool tendency should be facilitated by robust system response to selection.

In order of decreasing effect on replication (and thereby on entry into a selective Darwinian era), the important pool reactions are (compare Fig. 1, Fig. 4) as follows:

Nucleotide A and B dissociation constants (K_D) > substrate spike size (**spikesize**) > molecular lifetimes (k_{decay}^{-1}) > mean spike frequency (**freq**) > templated reaction rate (k_{template}) > nucleotide pairing rate (k_{bind}) > untemplated reaction rate ($k_{\text{notemplate}}$).

With these kinetic thresholds in hand, measured values for imidazolide-activated nucleotide behavior already published can be seen to approximate the calculated requirements. A major conclusion is, therefore, by using sporadic geochemical sources of varied, dilute, unstable activated 5' nucleotides reacting via already-known homogeneous chemical kinetics, a sporadically fed pool can readily become a cofactor-like system(s) that evolve(s).

The sporadically fed pool has a plausible early Earth context

Earth in the era of life's origin is arguably not hot but rather a heterogeneous ice ball. After the Moon-forming impact at about 4.5 gigayears ago, a very hot planet probably cooled in a few million years (Sleep *et al.*, 2001). Whereupon, given the lowered luminosity of the early Sun prior to its entry onto the stellar main sequence ($\approx 70\%$ of present luminosity; Sagan and Mullen, 1972), Earth could readily have descended to subfreezing temperatures (Zahnle, 2006). An ancient icy Earth was locally liquefied by bombardment by intersecting planetesimals (Bada *et al.*, 1994; Valley *et al.*, 2002), and a "cool early Earth" was ultimately thawed by a $\text{CO}_2/\text{H}_2\text{O}$ greenhouse effect (Sagan and Mullen, 1972; Bada *et al.*, 1994; Zahnle, 2006). Recent zircon isotopic data are consistent with clement conditions through this era, suggesting continental crust in contact with liquid water back to at least 4.3–4.4 gigayears (Mojzsis *et al.*, 2001; Harrison *et al.*, 2005).

Further, not only is an icy pool plausible on early Earth but in a surprising variety of other Solar System locales: for example, within Jupiter's satellite Europa (Pappalardo, 1999) and in the subsurface of saturnian satellite Enceladus (Postberg *et al.*, 2011). Thus sporadically fed pool events arguably bear on the breadth of life's distribution.

Cold water and ice are excellent media for RNA chemistry. Cold facilitates nucleotide pairing, and freezing facilitates reactions by concentrating solutes in the last-to-freeze eutectic phase (Kanavarioti *et al.*, 2001) and by stabilizing RNA structures disrupted by higher temperatures (Vlassov *et al.*, 2004). Chemical RNA synthesis is accordingly facilitated: templated addition of all four chemically activated ribonucleotides to a primer has been observed only in the cold (Deck *et al.*, 2011). Moreover, the eutectic phase itself forms a fine arboreal network that might have served to compartmentalize primordial biochemistry for evolution (Attwater *et al.*, 2010). In fact, the sporadically fed icy pool has successive access to both the concentrated reactants of the eutectic and the homogeneous effects of the pool. Alternation between liquid and eutectic deserves more investigation than it has received.

Further, a watery pool on glacier ice is a surprisingly benign environment for ribonucleotides. Modern Antarctic summer ponds, in areas free of endogenous biota, have a mean pH of 7.8, with 11 mM Na⁺, 1.1 mM Mg²⁺, and 0.25 mM Ca²⁺ from aerosol and liquid transport of dilute sea salts (de Mora *et al.*, 1994).

And not only is ice physically and chemically compatible with RNA chemistry, but seasonal and/or impact melting supplies a plausible valving mechanism to orchestrate sporadic flows of reactants, or transitions between the eutectic and solution. Notably, seasonal and impact flows would not be randomly timed but would be synchronized, likely elevating A-B yields over those calculated here.

Distribution of pool contents to nearby neighbors via melting is also an apt way of testing novel mixtures of pool contents. The potential of a population of pools is definitely beyond the scope of this manuscript, but production of something that depresses the freezing point of ice might be selected. In a sporadically freezing environment, this would lengthen the mean life of a pool at both thawing and freezing, and also catalyze pool fusion. An icy pool, that is, will exert its own native selection logic.

The broad consistency of an evolving sporadically fed RNA pool with an icy container (for example, compared to a rocky environment) is another rationale for minimal rates in calculations, to retain compatibility with lowered temperatures.

The posited pool product has a plausible position in biochemical history

A pool's reactive cofactor-like dinucleotide is readily connected to the later biochemical record (Yarus, 2011a), including modern organisms. *Bona fide* cofactors can be synthesized on oligomeric RNA by ribozymes (Huang *et al.*, 2000) or incorporated into RNAs using ribozyme activity (Jadhav and Yarus, 2002). They are useful biochemical reactants, once RNA-linked. For example, ribozymic synthesis of RNA-bound acetyl-CoA has been characterized (Jadhav and Yarus, 2002). Thus the chemical phenotype proposed

here exists and could be transmitted to a later RNA world composed of longer ribozymes. In turn, reactive modern 5'-5' dinucleotide cofactors used by proteins have long been cited as evidence of an ancestral generation of RNA enzymes, whose active groups were the progenitors of modern dinucleotide enzyme cofactors (White, 1976). Consequently, hypothetical pool events are plausibly continuously connected to modern biochemistry.

The sporadically fed pool has a plausible place before the big tree

Carl Woese has proposed (Woese, 2000) that there was an extended era before vertical inheritance was possible, recorded in the long, unbranched lines at the base of the big rRNA tree for Earthly life. This era supported evolution (*e.g.*, of the genetic code) in a population dominated by horizontal gene transfer. This is readily consistent with an ancestry in a sporadically fed pool, which readily divides and distributes its active molecules randomly but is too primitive to support reproducible vertical inheritance.

Epistemology

I conclude with comment on what is shown. A particular route to Darwinian biomolecules on Earth is not proven, but a credible IDA via small oligoribonucleotides now seems a likely hypothesis. Certainly, many potential problems—inconstant supplies of unstable precursors at low peak concentrations, or availability of only minimal catalysis—do not prove particularly challenging in the context of a realistic kinetic analysis. Further, a pool easily favors replication and efficient selection for a more catalytically effective product, the dimer, taking advantage of natural variance in molecular reactivities at almost any elementary step (Figs. 1 and 4). Known ribonucleotide reactions thus appear sufficient for robust progress from undirected beginnings (near zeroth order in A-B, leftward in Fig. 4 panels) to a system structured by Darwinian change (near first order in A-B, rightward in Fig. 4 panels). Nonetheless, particular data require further investigation. Most urgently, stabilities, pairing, and reaction rates for varied, small, 3'-5', 2'-5', and 5'-5' oligonucleotides in eutectic ice and cold water are crucially influential and insufficiently known. A primordial leaving group Ψ , with rates similar to phosphorimidazolides, remains to be specified. Nevertheless, calculations find no bar to the appearance of an IDA via minimally catalyzed nucleotide chemistry at lowered temperatures. Such an origin for terran life, as for Darwinian biology elsewhere, is not unimaginable. Instead, it is entirely feasible to discuss its necessities.

Online Supplements

Two supplementary files (*see* Supplementary Information at www.liebertonline.com/ast) are available that clarify these calculations:

Numerical Results – full Figure 4 quantitation
Berkeley Madonna Simulation Code – the runnable Figure 1 model.

Abbreviation

IDA, Initial Darwinian Ancestor.

References

- Agris, P.F., Vendeix, F.A., and Graham, W.D. (2007) tRNA's wobble decoding of the genome: 40 years of modification. *J Mol Biol* 366:1–13.
- Attwater, J., Wochner, A., Pinheiro, V.B., Coulson, A., and Holliger, P. (2010) Ice as a protocellular medium for RNA replication. *Nat Commun* 1:1–9.
- Bada, J.L., Bigham, C., and Miller, S.L. (1994) Impact melting of frozen oceans on the early Earth: implications for the origin of life. *Proc Natl Acad Sci USA* 91:1248–1250.
- Darwin, C. (1859) *The Origin of Species by Means of Natural Selection*, John Murray, London.
- de Mora, S.J., Whitehead, R.F., and Gregory, M. (1994) The chemical composition of glacial melt water ponds and streams on the McMurdo Ice Shelf, Antarctica. *Antarct Sci* 6:17–27.
- Deck, C., Jauker, M., and Richert, C. (2011) Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nat Chem* 3:603–608.
- Feller, W. (1957) Chapter X—Laws of large numbers. In *An Introduction to Probability Theory and Its Applications*, John Wiley & Sons, Hoboken, NJ, pp 228–231.
- Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T., and Turner, D.H. (1986) Improved free-energy parameters for predictions of RNA duplex stability. *Proc Natl Acad Sci USA* 83:9373–9377.
- Harrison, T.M., Blichert-Toft, J., Muller, W., Albarede, F., Holden, P., and Mojzsis, S.J. (2005) Heterogeneous Hadean hafnium: evidence of continental crust at 4.4 to 4.5 Ga. *Science* 310:1947–1950.
- Huang, F., Bugg, C.W., and Yarus, M. (2000) RNA-catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN. *Biochemistry* 39:15548–15555.
- Illangasekare, M. and Yarus, M. (1999) A tiny RNA that catalyzes both aminoacyl-RNA and peptidyl-RNA synthesis. *RNA* 5:1482–1489.
- Illangasekare, M. and Yarus, M. (2012) Small aminoacyl transfer centers at GU within a larger RNA. *RNA Biol* 9:59–66.
- Inoue, T. and Orgel, L.E. (1983) A nonenzymatic RNA polymerase model. *Science* 219:859–862.
- Jadhav, V.R. and Yarus, M. (2002) Acyl-CoAs from coenzyme ribozymes. *Biochemistry* 41:723–729.
- Kanavarioti, A., Rosenbach, M.T., and Hurley, T.B. (1992) Nucleotides as nucleophiles: reactions of nucleotides with phosphoimidazole activated guanosine. *Orig Life Evol Biosph* 21:199–217.
- Kanavarioti, A., Bernasconi, C.F., and Baird, E.E. (1998) Effects of monomer and template concentration on the kinetics of non-enzymatic template-directed oligoguanylate synthesis. *J Am Chem Soc* 120:8575–8581.
- Kanavarioti, A., Baird, E.E., Hurley, T.B., Carruthers, J.A., and Gangopadhyay, S. (1999) Unique catalysis and regioselectivity observed in the poly(C)-directed RNA dimer formation from 2-MeImpG: kinetic analysis as a function of monomer and polymer concentration. *J Org Chem* 64:8323–8333.
- Kanavarioti, A., Monnard, P.A., and Deamer, D.W. (2001) Eutectic phases in ice facilitate nonenzymatic nucleic acid synthesis. *Astrobiology* 1:271–281.
- Kennedy, R., Lladser, M.E., Yarus, M., and Knight, R. (2008) Information, probability, and the abundance of the simplest RNA active sites. *Front Biosci* 13:6060–6071.
- Labuda, D., Striker, G., Grosjean, H., and Porschke, D. (1985) Mechanism of codon recognition by transfer RNA studied with oligonucleotides larger than triplets. *Nucleic Acids Res* 13:3667–3683.
- Larralde, R., Robertson, M.P., and Miller, S.L. (1995) Rates of decomposition of ribose and other sugars: implications for chemical evolution. *Proc Natl Acad Sci USA* 92:8158–8160.
- Levy, M. and Miller, S.L. (1998) The stability of the RNA bases: implications for the origin of life. *Proc Natl Acad Sci USA* 95:7933–7938.
- Lifson, S. and Lifson, H. (1999) A model of prebiotic replication: survival of the fittest versus extinction of the unfittest. *J Theor Biol* 199:425–433.
- Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature* 362:709–715.
- Lohrmann, R. (1977) Formation of nucleoside 5'-phosphoramidates under potentially pre-biological conditions. *J Mol Evol* 10:137–154.
- Mojzsis, S.J., Harrison, T.M., and Pidgeon, R.T. (2001) Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* 409:178–181.
- Oro, J. (1961) Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive Earth conditions. *Nature* 191:1193–1194.
- Pace, N.R., Sapp, J., and Goldenfeld, N. (2012) Phylogeny and beyond: scientific, historical, and conceptual significance of the first tree of life. *Proc Natl Acad Sci USA* 109:1011–1018.
- Pappalardo, R., Belton, M.J.S., Breneman, H.H., Carr, M.H., Chapman, C.R., Collins, G.C., Denk, T., Fagents, S., Geissler, P.E., Giese, B., Greeley, R., Greenberg, R., Head, J.W., Helfenstein, P., Hoppa, G., Kadel, S.D., Klaasen, K.P., Klemaszewski, J.E., Magee, K., McEwen, A.S., Moore, J.M., Moore, W.B., Neukum, G., Phillips, C.B., Prockter, L.M., Schubert, G., Senske, D.A., Sullivan, R.J., Tufts, B.R., Turtle, E.P., Wagner, R., and Williams, K.K. (1999) Does Europa have a subsurface ocean? Evaluation of the geological evidence. *J Geophys Res* 104:24015–24055.
- Postberg, F., Schmidt, J., Hillier, J., Kempf, S., and Srama, R. (2011) A salt-water reservoir as the source of a compositionally stratified plume on Enceladus. *Nature* 474:620–622.
- Powner, M.W., Gerland, B., and Sutherland, J.D. (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459:239–242.
- Rohatgi, R., Bartel, D.P., and Szostak, J.W. (1996) Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. *J Am Chem Soc* 118:3332–3339.
- Rosenbrock, H.H. (1963) Some general implicit processes for the numerical solution of differential equations. *The Computer Journal* 5:329–330.
- Sagan, C. and Mullen, G. (1972) Earth and Mars: evolution of atmospheres and surface temperatures. *Science* 177:52–56.
- Sawai, H. and Orgel, L.E. (1975) Oligonucleotide synthesis catalyzed by the zinc ion. *J Am Chem Soc* 97:3532–3533.
- Scheuring, I. and Szathmari, E. (2001) Survival of replicators with parabolic growth tendency and exponential decay. *J Theor Biol* 212:99–105.
- Schoenberg, D.R. and Maquat, L.E. (2009) Re-capping the message. *Trends Biochem Sci* 34:435–442.
- Sievers, D. and von Kiedrowski, G. (1994) Self-replication of complementary nucleotide-based oligomers. *Nature* 369:221–224.
- Sleep, N.H., Zahnle, K., and Neuhoff, P.S. (2001) Initiation of clement surface conditions on the earliest Earth. *Proc Natl Acad Sci USA* 98:3666–3672.

- Soukup, G.A. and Breaker, R.R. (1999) Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* 5:1308–1325.
- Sulston, J., Lohrmann, R., Orgel, L.E., and Miles, H.T. (1968) Nonenzymatic synthesis of oligoadenylates on a polyuridylic acid template. *Proc Natl Acad Sci USA* 59: 726–733.
- Szathmáry, E. and Maynard Smith, J. (1997) From replicators to reproducers: the first major transitions leading to life. *J Theor Biol* 187:555–571.
- Szostak, J.W., Bartel, D.P., and Luisi, P.L. (2001) Synthesizing life. *Nature* 409:387–390.
- Valley, J.W., Peck, W.H., King, E.M., and Wilde, S.A. (2002) A cool early Earth. *Geology* 30:351–354.
- Vlassov, A., Johnston, B., Landweber, L., and Kazakov, S. (2004) Ligation activity of fragmented ribozymes in frozen solution: implications for the RNA world. *Nucleic Acids Res* 32:2966–2974.
- von Kiedrowski, G. (1986) A self-replicating hexadeoxynucleotide. *Angew Chem Int Ed Engl* 25:932–935.
- Westheimer, F.H. (1987) Why nature chose phosphates. *Science* 235:1173–1178.
- White, H.B., III. (1976) Coenzymes as fossils of an earlier metabolic state. *J Mol Evol* 7:101–104.
- Woese, C.R. (2000) Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci USA* 97:8392–8396.
- Yarus, M. (2011a) Getting past the RNA world: the initial Darwinian ancestor. In *RNA Worlds: From Life's Origins to Diversity in Gene Regulation*, edited by J.F. Atkins, T.R. Cech, and R.F.S. Gesteland, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 43–50.
- Yarus, M. (2011b) The meaning of a minuscule ribozyme. *Philos Trans R Soc Lond B Biol Sci* 366:2902–2909.
- Yarus, M., Caporaso, J.G., and Knight, R. (2005) Origins of the genetic code: the escaped triplet theory. *Annu Rev Biochem* 74:179–198.
- Zahnle, K.J. (2006) Earth's earliest atmosphere. *Elements* 2:217–222.

Address correspondence to:

Michael Yarus
Molecular, Cellular and Developmental Biology
University of Colorado
Boulder, CO 80309-0347

E-mail: yarus@stripe.colorado.edu

Submitted 7 April 2012

Accepted 29 June 2012